

METHOD FOR COMPUTING MODELS BASED ON
ATTRIBUTES SELECTED BY ENTROPY
ABSTRACT OF THE DISCLOSURE

5
Sub A 37
10
15
20
25
Attributes of a data set to be employed in generating a predictive model are analyzed based on entropy, chi-square, or similar statistical measure. A target group of samples exhibiting one or more desired attributes is identified, then remaining attribute values for the target group are compared to corresponding attribute values for the whole sample population. A subset of all available attributes is then selected from those attributes which exhibit, when comparing attribute values of target group samples to attribute values for the whole sample population, the greatest relative difference or divergence. That is, an attribute for which the target group samples exhibit, for example, only two of all possible values is selected in preference to an attribute for which the target group samples exhibit three or more of the possible values. This subset is employed to generate the predictive model. Efficiency in generating the predictive model is improved, since fewer attributes are employed and less computational resources are required. Accuracy of the resulting predictive model is also improved since attributes potentially skewing the sample population in a manner least related to the desired attribute are eliminated from consideration in developing the model.